SOLUTIONS TO MIDTERM

**Problem 1** Consider the online paging problem with cache size $k$ and $n$ total memory items, with $1 \leq k < n$, and suppose the requests are random and independent, with each request being uniformly distributed over the $n$ memory items.
(a) What is the miss rate (misses per request in the long run) for any online paging algorithm?
(b) Recall that the MARKER algorithm works in rounds. What is the expected number of misses of the MARKER algorithm per round? (Solve independently of (a)).
(c) What is the expected number of requests per round of the MARKER algorithm? (Solve independently of (a)).
(d) How are the answers to (a), (b), and (c) related? (This does not require knowledge of the MARKER algorithm.)

SOLUTION
(a) Each request is a miss with probability $1 - \frac{k}{n}$, independently of previous requests. So the miss rate is $1 - \frac{k}{n}$ in the long run with probability one, by the law of large numbers.
(b) The first request of a round is always a miss. After $i$ distinct requests have been made in a round, the next distinct request to arrive in the round results in a miss with probability $\frac{n-k}{n-i}$, for $1 \leq i \leq k-1$. Thus, the expected number of misses per round is $1 + \sum_{i=1}^{k-1} \frac{n-k}{n-i} = \sum_{i=1}^{k} \frac{n-k}{n-i}$.
(c) The mean length of a round is $\sum_{i=1}^{k} \frac{n}{n-i}$, where the $i$th term is the mean number of requests including or after the $i$th distinct request in the round, but strictly before the $i+1$th distinct.
(d) By basic renewal theory or the law of large numbers, the answer to (a) is the anwer to (b) divided by the answer to (c).

**Problem 2** Recall that in each step of Luby's LT decoding process an output symbol with reduced degree one (i.e. a symbol from the gross ripple) is selected and processed. It is decoded, its corresponding input symbol is subtracted from the other output symbols containing it, and the degrees of those output symbols are reduced by one. The decoding process ends when the gross ripple is empty at the end of some step. Let $A_o$ be the set of input symbols that are not decoded after the process ends. Show that $A_o$ is completely determined by the set of output symbols initially given to the decoder. That is, $A_o$ does not depend on which output symbol is chosen from the ripple at each step. (Hint: It may be helpful to characterize $A_o$ in a way that is independent of the choices made by the decoding algorithm.)

SOLUTION
Call a set of input symbols $B$ a blocking set if no output symbol is connected to exactly one symbol in $B$. If $B$ is a blocking set then the LT decoding process will not decode any symbols in $B$, no matter what order symbols are processed in. When the LT decoding process ends, the set of undecoded symbols is a maximal blocking set, meaning that it is a blocking set and no larger blocking set contains it. Observe that if $B_1$ and $B_2$ are both blocking sets, then $B_1 \cup B_2$ is also a blocking set. Therefore, there exists a unique maximal blocking set. Therefore, $A_o$ is the unique maximal blocking set.

A perhaps simpler approach is to show by induction on $k$ that, for any two decoding processes, all of the first $k$ symbols decoded by the first process are eventually decodable by the second decoding process.

**Problem 3** Given $1 \leq d < n$, let $U = \{u_1, \ldots, u_n\}$ and $V = \{v_1, \ldots, v_n\}$ be disjoint sets of cardinality $n$, and let $G$ be a bipartite random graph with vertex set $U \cup V$, such that if $V_i$ denotes the set of neighbors of $u_i$, then $V_1, \ldots, V_n$ are independent, and each is uniformly distributed over the set of all $\binom{n}{d}$ subsets of $V$ of cardinality $d$. A matching for $G$ is a subset of edges $M$ such that no two edges in $M$ have a common vertex. Let $Z$ denote the maximum of the cardinalities of the matchings for $G$.
(a) Find bounds $a$ and $b$, with $0 < a \leq b < n$, so that $a \leq E[Z] \leq b$.
(b) Give an upper bound on $P\{|Z - E[Z]| \geq \gamma\sqrt{n}\}$, for $\gamma > 0$, showing that for fixed $d$, the distribution of $Z$ is concentrated about its mean as $n \to \infty$.
(c) Suggest a greedy algorithm (similar to the LT process) for finding a large cardinality matching.

SOLUTION
(a) Many bounds are possible. Generally the tighter bounds are more complex. If $d = 1$, then every vertex in $V$ of degree one or more can be included in a single matching, and such a matching is a maximum matching with expected size $n(1 - (1 - \frac{1}{n})^n) \geq n(1 - e^{-1})$. The cardinality of a maximum matching cannot decrease as more edges are added, so $a = n(1 - e^{-1})$ is a lower bound for any $d$ and $n$. An upper bound is given by the expected number of vertices of $V$ with degree at least one, so $b = n(1 - (1 - \frac{d}{n})^n) \approx n(1 - e^{-d})$ works.
(b) The variable $Z$ can be expressed as $Z = F(V_1, \ldots, V_n)$, where $V_i$ is the set of neighbors of $u_i$, and $F$ satisfies the Lipschitz condition with constant 1. We are thus considering a martingale associated with $Z$ by exposing the vertices of $U$ one at a time. The (improved version of the) Azuma-Hoeffding inequality yields that $P\{|Z - E[Z]| \geq \gamma\sqrt{n}\} \leq 2e^{-2\gamma^2}$.
(c) Process the vertices of $U$ one at a time. Select an unprocessed vertex in $U$. If the vertex has positive degree, select one of the edges associated with it and add it to the matching. Remove edges from the graph that become ineligible for the matching, and reduce the degrees of the unprocessed vertices correspondingly. (This algorithm can be improved by selecting the next vertex in $U$ from among those with smallest reduced degree, and also selecting outgoing edges with endpoint in $V$ having the smallest reduced degree in $V$, or both. If every matched edge has an endpoint with reduced degree one at the time of matching, the resulting matching has maximum cardinality. )